

Rupali Roy (UTID : rur73)

Paper: SUN database Large-scale Scene Recognition from Abbey to Zoo

Q1. How was the taxonomy created?

Ans. The taxonomy for scene classification was created by using around 70,000 terms from WordNet that were available on the tiny image dataset. The terms that were included described the scenes, places and environments. Most of the terms like *outdoors*, didn't give any specific visual identity were excluded along with non-navigable scenes and vehicles. To get accurate environmental category high tolerance level was maintained for vocabulary terms. After combining all the terms from the WordNet and the terms that gave some visual identity, the final dataset contained around 899 categories. (Xiao, 2010, p.3)

Q2. What was the design of the crowdsourcing task used to assign images to a scene category?

Ans. The 397 environmental categories of the SUN database are divided into 3 tree level hierarchy and the crowd workers navigate through it to arrive at any specific scene type. Such 3-level hierarchies help the crowd workers to easily label the categories that were otherwise unfamiliar to most of them. The accuracy for the scene classification is done by using Amazon Mechanical Turk. (Xiao, 2010, p.4)

Q3. What quality control methods were used to ensure high quality image labels were collected from crowd workers?

Ans. To ensure high quality image labels the categories are first divided into 3 tree level hierarchy. It then considers 20 distinct test scenes for all the 397 categories which results in 7940 Human Intelligence Task. As mentioned in the paper the average time workers took is around 61 seconds for each HIT task with leaf level accuracy of 58.6%. The paper also mentions about "good" and "bad" workers depending upon their accuracy level. The good workers have 95% of accuracy at leaf level and 68.5% accuracy at leaf level. (Xiao, 2010, p.4)

Q4. What methods were used to demonstrate the advantage of the newly proposed dataset over existing datasets?

Ans. Several state of art features like GIST, HOG 2*2, Dense SIFT, LBP, Sparse SIFT Algorithms, SSIM, Tiny Images, Line Features, Texton Histograms, Color Histograms, Geometric Probability Map and Geometry Specific Histograms were used to demonstrate the advantage of the newly proposed dataset over the existing datasets. (Xiao, 2010, p.5)

LSUN:Construction of a Large-Scale Image Dataset using Deep Learning with Humans in the Loop

1. How was the taxonomy created?

Ans. The taxonomy was created using Deep Learning Algorithms for image classification and with the help of humans to label an image. Various classification models are considered for effective image classification. Multi Layer Perceptron along with pretrained GoogLeNet is used. The image sampling is done using multiple iteration that result in positive and negative image samples. (Yu,2016,p.3) (Yu,2016,p.4)

2. What was the design of the crowdsourcing task used to assign images to a scene category?

Ans. Amazon Mechanical Turk is the crowdsourcing platform used to obtain high quality image annotation for the image dataset. To help and ease the task of humans to label data accurately an interface is designed that allows a person to see one image at a time along with the definition of the category. It facilitates human to see thumbnails of the pervious and next image that are in queue for labeling. To increase the accuracy for efficient labeling of images the crowd workers are presented with an instruction page along with some category specific examples. To get high quality annotated image, Quality Control is enforced by collecting image from two crowd workers and keeping only the doubly confirmed labels.(Yu,2016,p.5)

3. What quality control methods were used to ensure high quality image labels were collected from crowd workers?

Ans. To ensure that high quality image labels were collected from crowd workers the concept of redundant labeling and enforced instruction. For each annotated image, data from two crowd workers is collected and only doubly confirmed labels are kept. The quality control method used is described as: If we have a set of around 150 images then before the humans start labeling the images, we append 55 more images for quality control purpose; now the total images that needs to be labeled are 205 out of which 150 are actual images. Out of the 55 images that are used for quality control, 15 images are to test if crowd workers understand the instructions properly and out of the remaining 40 images, 20 are online images whose quality is pre checked and 20 other images are checked after HIT is submitted. .(Yu,2016,p.6)

4. What methods were used to demonstrate the advantage of the newly proposed dataset over existing datasets?

Ans. The author uses the below mentioned characteristics or methods to demonstrate the advantage of the LSUN database over other existing datasets are:

1. Dataset Statistics
2. Label Precision
3. Effort Amplification
4. Impact on Model Performance
5. Learned Image Representation (Yu,2016,p.6) (Yu,2016,p.7)

5. In addition, please submit one "discussion point" in total (not per paper). This can be in the form of a question, critique, connection to other readings, or plausible future work that you think is interesting to investigate in greater detail in class. You can choose to have this point be about any of the scene classification papers (i.e., SUN, Places2, and LSUN). The discussion point may be about one paper or can compare and contrast both papers you chose to read. In addition, the discussion point can address the proposed idea, methods, experimental design, and analysis of results. This should be about 1-2 sentences long.

Ans. For The LSUN database it's mentioned that we use multi-layer perceptron with two hidden layers. I would like to discuss how MLP is used for this specific dataset and how we decide the number of hidden layers to get accurate data in MLP . (Yu,2016,p.4)